SUBMITTED TO APJ, OCTOBER 11, 2012
Preprint typeset using LaTeX style emulateapj v. 5/2/11

OPTIMAL CORRELATION ESTIMATORS FOR QUANTIZED SIGNALS

M. D. JOHNSON, H. H. CHOU, C. R. GWINN
Department of Physics, University of California, Santa Barbara, California 93106, USA
Submitted to ApJ, October 11, 2012

ABSTRACT

Using a maximum-likelihood criterion, we derive optimal correlation strategies for signals with and without digitization. We assume that the signals are drawn from zero-mean Gaussian distributions, as is expected in radio-astronomical applications, and we present correlation estimators both with and without *a priori* knowledge of the signal variances. We demonstrate that traditional estimators of correlation, which rely on averaging products, exhibit large and paradoxical noise when the correlation is strong. However, we also show that these estimators are fully optimal in the limit of vanishing correlation. We calculate the bias and noise in each of these estimators and discuss their suitability for implementation in modern digital correlators.

Subject headings: methods: data analysis – methods: statistical – techniques: interferometric

1. INTRODUCTION

Because astrophysical sources emit Gaussian noise, the information in astrophysical observations lies in signal covariances. These are familiar as power spectra and cross spectra (see, for example, Thompson et al. 2001). The estimation of these covariances is subject to bias and noise, and techniques to minimize both are therefore fundamental to radio astronomy.

This estimation is complicated by the typically aggressive quantization of the received signal. Even for next-generation phased arrays, such as the Square-Kilometer Array, the cost of signal transmission necessitates low-bit quantization (Dewdney et al. 2009). This procedure distorts the spectrum but preserves much of the underlying statistical information. In fact, several authors have noted the ability of quantization to *improve* estimates of correlation $\rho \in [-1, 1]$, especially for strong correlation $|\rho| \to 1$ (Gwinn 2004). Indeed, Cole (1968) found that the standard two-level correlation scheme has lower noise than four- and six-level schemes in this limit, and that all of these estimates have lower noise than the correlation estimates for unquantized signals. We demonstrate that this paradoxical behavior arises from two causes: comparisons with unquantized correlation estimates are incomplete, and typical quantization schemes are not optimal.

To amend these deficiencies, we present a correlation estimator for unquantized data that is appropriately suited to define a quantization efficiency, and we derive optimal correlation estimators for quantized signals via a maximum-likelihood criterion. With the recent advent of digital correlators in radio astronomy, such as DiFX (Deller et al. 2007), implementing these techniques is straightforward.

1.1. *Terminology and Notation*

In spite of the many treatments of quantized correlation, no standard terminology has been adopted, so we first outline some basic assumptions and definitions. Throughout this work, we use $\{x_i, y_i\}$ to designate sets of pairs independently drawn from a zero-mean bivariate Gaussian distribution. For simplicity, we will assume that the standard deviations $\sigma_{\rm x}$ and $\sigma_{\rm y}$ are unity.

We denote ensemble averages by unsubscripted angular brackets $\langle \ldots \rangle$. We will make use of the *correlation* $\rho \equiv \langle xy \rangle / (\sigma_{\rm x}\sigma_{\rm y})$ and the *covariance* $\langle xy \rangle$.

We denote finite averages, over a sample of N points, by subscripted angular brackets $\langle \ldots \rangle_N$. For example, we frequently use the *sample covariance* $r_\infty \equiv \langle xy \rangle_N = N^{-1} \sum_{i=1}^N x_i y_i$. We also use this terminology to refer to an average product after quantization. Because most applications of correlation in radio astronomy involve many samples N, we focus on the large–N regime.

Our work focuses on estimators of ρ, given a set of N samples $\{x_i, y_i\}$, possibly after quantization. We use the variable r, with subscripted identifiers, to indicate such estimators. Finally, we generically use $P(\ldots)$ to denote a probability density function (PDF) with respect to the given variables and parameters.

1.2. *Relation to Previous Work*

Previous analyses of quantized correlation have assumed that the correlation should be estimated via a form of sample covariance for the quantized signals; they have then optimized the performance of the correlation by choosing an appropriate quantization scheme. Furthermore, these efforts generally focus on the small correlation regime: $|\rho| \ll 1$.

For example, Jenet & Anderson (1998) provide an approximate prescription for correcting the bias from quantization in sample covariance. However, this prescription still suffers from severely sub-optimal performance when $\rho \neq 0$, in terms of the noise.

In contrast, we provide a new mechanism for estimating correlation and demonstrate that it provides the lowest RMS error of *any* post-quantization correlation strategy for a large number of samples. We also demonstrate this this strategy is equivalent to traditional approaches as $\rho \to 0$, and we give a rigorous justification for the optimal weights that are typically quoted.

1.3. *Outline of Paper*

michaeltdh@physics.ucsb.edu,cgwinn@physics.ucsb.edu

In §2, we briefly review the basic mathematical framework of parameter estimation theory, and we define the sense in which a particular strategy can be "optimal." Then, in §3, we consider the case of unquantized signals and present the corresponding optimal estimators for correlation. Next, in §4, we summarize the details of the quantization procedure, outline the traditional correlation estimators via sample covariance, and derive the maximum-likelihood estimate of correlation for quantized signals. In §5, we give specific examples for common quantization schemes, and compare the performance of the maximum-likelihood estimate to that of traditional estimates. Then, in §6, we demonstrate the critical property that traditional correlation schemes are optimal for small $|\rho|$. Finally, in §7, we summarize our findings and discuss the possibilities for implementation.

2. MATHEMATICAL BACKGROUND

We begin by reviewing some essential concepts and terminology in parameter estimation theory. For a comprehensive discussion of these ideas with a rigorous description of the assumptions and constraints, see Kendall & Stuart (1979) or Lehmann & Casella (1998).

2.1. *Optimal Estimators and Maximum Likelihood*

We first ascribe a precise meaning to the term "optimal" estimator. For this purpose, we must consider both the bias and noise in an estimator. We seek estimates of correlation that converge to the exact correlation as $N \to \infty$; such estimates are said to be *consistent.* We refer to a consistent estimator with the minimum noise (i.e. the minimum mean squared error) as the optimal estimator.

If the parameters to be estimated correspond to a known class of distributions, then a particularly simple estimator can be defined. Namely, consider a set of observations $\mathbf{x}$ drawn from a distribution that is specified by a set of parameters $\boldsymbol{\theta}_0$. One parameter estimation strategy determines the parameters which maximize the likelihood function $\mathcal{L}(\boldsymbol{\theta}|\mathbf{x})$, defined as the probability of sampling $\mathbf{x}$ given the distribution specified by $\boldsymbol{\theta}$. If $\mathcal{L}$ has a unique maximum at some $\hat{\boldsymbol{\theta}}_{\rm ML}$, then this point is defined to be the *maximum-likelihood estimator* (MLE) of $\boldsymbol{\theta}_0$ for the sampled points $\mathbf{x}$.

Often, the sample data $\mathbf{x}$ can be greatly reduced to some simplified statistic $\mathbf{T}(\mathbf{x})$, which carries all the information related to the parameters $\boldsymbol{\theta}_0$. In this case, $\mathbf{T}(\mathbf{x})$ is said to be a *sufficient* statistic for $\boldsymbol{\theta}_0$. For example, if samples are drawn from a normal distribution with known variance but unknown mean, then the sample mean is a sufficient statistic for the mean. The *factorization criterion* states that a necessary and sufficient condition for $\mathbf{T}(\mathbf{x})$ to be sufficient for a family of distributions parametrized by $\boldsymbol{\theta}_0$ is that there exist non-negative functions g and h such that $P(\mathbf{x};\boldsymbol{\theta}_0) = g[\mathbf{T}(\mathbf{x});\boldsymbol{\theta}_0]h(\mathbf{x})$.

Under weak regularity conditions, the likelihood function also determines the minimum noise that *any* unbiased estimator can achieve. This minimum, the Cramér-Rao bound (CRB), can be expressed in terms of derivatives of $\mathcal{L}$. For example, the minimum variance of any unbiased estimator $\hat{\theta}$ of a single parameter θ_0 is the inverse of the *Fisher information*, and can be written

$$\left\langle \delta\hat{\theta}^2 \right\rangle \geq \left\langle \left(\left.\frac{\partial \ln \mathcal{L}(\mathbf{x};\theta)}{\partial \theta}\right]_{\theta=\theta_0} \right)^2 \right\rangle^{-1} \equiv \delta\hat{\theta}^2_{\rm CR}. \quad (1)$$

Here, $\langle \ldots \rangle$ denotes an ensemble average over sets of measurements $\mathbf{x}$. An unbiased estimator with noise that matches the CRB is said to be *efficient*.

Under general conditions, the MLE is both consistent and asymptotically (as $N \to \infty$) efficient. In the present work, we present the MLE of correlation for both unquantized and quantized signals, and we compare these correlation strategies with traditional schemes.

2.2. *Distribution of Correlated Gaussian Variables*

Astrophysical observations measure zero-mean, Gaussian noise. Under rather broad assumptions, pairs of such samples $\{x, y\}$ are drawn from a bivariate Gaussian distribution. In addition to the respective standard deviations, $\sigma_{\rm x} \equiv \sqrt{\langle x^2 \rangle}$ and $\sigma_{\rm y} \equiv \sqrt{\langle y^2 \rangle}$, this distribution depends on the correlation $\rho \equiv \langle xy \rangle/(\sigma_{\rm x}\sigma_{\rm y}) \in [-1, 1]$. Because our present emphasis is correlation, we assume that $\sigma_{\rm x} = \sigma_{\rm y} = 1$, in which case the PDF is given by

$$P(x, y; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x^2 + y^2 - 2\rho xy}{2\,(1-\rho^2)} \right]. \quad (2)$$

For small $|\rho|$, this distribution takes the following approximate form:

$$P(x, y; \rho) \approx \frac{1}{2\pi}\,(1 + \rho xy)\, e^{-\frac{1}{2}\left(x^2+y^2\right)}. \quad (3)$$

See Chapter 8 of Thompson et al. (2001) (hereafter TMS) for some additional representations and discussion.

3. CORRELATION ESTIMATORS FOR UNQUANTIZED SIGNALS

We now analyze several correlation estimators for unquantized signals. These estimators serve two relevant purposes: they provide a point of reference to ascribe an efficiency to a quantization scheme, and they suggest closed-form strategies for correlation estimates of quantized signals that have a large number of bits.

First, in §3.1, we consider the estimate of correlation via sample covariance, denoted r_∞. Next, in §3.2, we present Pearson's estimate of correlation, $r_{\rm p}$, which serves as the optimal estimator when no information about the signal is known. Last, in §3.3, we give details of the MLE of correlation when the signal variances are known, which we denote $r_{\rm q}$. Figure 1 compares the asymptotic noise in these three estimates, as given in the following sections.

3.1. *Correlation via Sample Covariance:* r_∞

The simplest estimate of correlation follows from the relationship between correlation and covariance. Namely, suppose that the means $\{\mu_{\rm x}, \mu_{\rm y}\}$ and standard deviations $\{\sigma_{\rm x}, \sigma_{\rm y}\}$ of the signals x_i and y_i are known. In this case, the signals may be standardized to have zero mean and unit variance. Their covariance is then equal to their correlation: $\langle xy \rangle = \rho$. This correspondence immediately suggests a simple estimator for the correlation: $r_\infty \equiv \langle xy \rangle_N = N^{-1} \sum_{i=1}^{N} x_i y_i$.

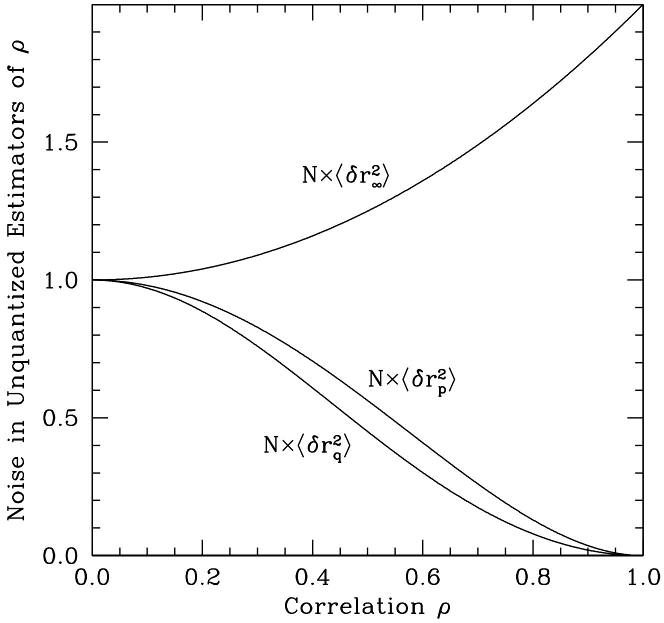

Figure 1. Asymptotic noise in estimators r_∞, r_p, and r_q of ρ for unquantized signals. Because the noise is a symmetric function of ρ, only the positive values are shown.

This estimator is unbiased and consistent but has large variance: $N\langle \delta r_\infty^2\rangle \equiv N\langle (r-\rho)^2\rangle = \left(1+\rho^2\right)$ (TMS; Eq. 8.13). This equation highlights a peculiar feature: r_∞ is the noisiest when the correlation is *strongest*.

3.2. *Optimal Correlation:* r_p

Many researchers have studied improved estimators of correlation (see Hotelling (1953) and Anderson (1996) for interesting perspectives). The most common estimator is known as "Pearson's r" and is given by

$$r_\mathrm{p} \equiv \frac{\left\langle \left(x-\langle x\rangle_N\right)\left(y-\langle y\rangle_N\right)\right\rangle_N}{\sqrt{\left\langle \left(x-\langle x\rangle_N\right)^2\right\rangle_N \left\langle \left(y-\langle y\rangle_N\right)^2\right\rangle_N}}. \quad (4)$$

In addition to being unbiased and consistent, r_p is asymptotically efficient. The asymptotic noise can be derived using the Fisher transformation (Fisher 1915, 1921): $\lim_{N\to\infty} N\left\langle \delta r_\mathrm{p}^2\right\rangle = \left(1-\rho^2\right)^2$, which is indeed the CRB.

Note that substituting the exact means and variances into r_p returns the original estimate r_∞ and, remarkably, *decreases* the quality of the estimate. As a simple example, three randomly generated samples with correlation $\rho = 0.999$ are $\{x_i\} = \{0.998, 1.712, -0.992\}, \{y_i\} = \{1.01, 2.01, -0.980\}$. In this case, we obtain estimates $r_\infty = 1.81$, $r_\mathrm{p} = 0.997$. Indeed, for perfect correlation, $\delta r_\mathrm{p}^2 = 0$, whereas $\delta r_\infty^2 = 2/N$. Pearson's estimate accounts for the sample variance, which contributes much of the noise in r_∞. However, for small correlation, $\rho \to 0$, the noise in r_∞ and r_p is identical (we further discuss this feature in §6).

Hence, when the correlation is large, simply averaging products poorly approximates the correlation relative to other schemes. Even if the exact variance is known, the sample variance must still be incorporated to optimally estimate the correlation.

3.3. *Optimal Correlation with Known Signal Variance:* r_q

Nevertheless, an exact knowledge of the variance can be used to effectively improve the estimate of correlation. In fact, this knowledge is generally assumed in radio astronomy. For example, automatic gain control usually sets the variances $\langle x^2\rangle = \langle y^2\rangle = 1$, and quantization schemes use the "known" variance to determine the appropriate level settings. Errors in the signal estimate are then a source of both bias and noise, so for non-stationary signals such as pulsars, the quantization weights must be dynamically adjusted (see Jenet & Anderson 1998). For a more complete discussion of quantization noise, see Gwinn (2004) and Gwinn (2006).

More generally, whenever the timescale of variation of ρ is shorter than that of variation in the standard deviation σ of x and y, there will be improved measures of correlation.

For example, if the standard deviations of x and y are known to be unity, then the MLE of correlation, denoted r_q, is determined by the real solution of

$$r_\infty\left(1+r_\mathrm{q}^2\right) - r_\mathrm{q}\left(\langle x^2\rangle_N + \langle y^2\rangle_N - \left(1-r_\mathrm{q}^2\right)\right) = 0. \quad (5)$$

In §A, we derive this result, give an approximate form for r_q, and demonstrate that the noise in this estimate achieves the CRB as $N\to\infty$, as expected for an MLE: $\lim_{N\to\infty} N\left\langle \delta r_\mathrm{q}^2\right\rangle = \left(1-\rho^2\right)^2/\left(1+\rho^2\right)$. The advantage of r_q relative to r_p increases with $|\rho|$, and gives a factor of two improvement in the estimator variance at high correlation. Moreover, the bias in r_q is $o\left(N^{-1}\right)$.

4. CORRELATION ESTIMATORS FOR QUANTIZED SIGNALS

In practice, data are digitized, which involves quantization according to a prescribed scheme. The merit of the quantization, reduction of data volume, must be carefully weighed against its drawback, degraded signal information.

We first review the details of quantization and the traditional estimators of correlation, which rely on the sample covariance after quantization. We then derive the MLEs of correlation for arbitrary quantization schemes and give expressions for the noise in these estimators.

4.1. *The Quantization Transfer Function*

The process of quantization maps each element in a time-series $x_i \in \mathbb{R}$ to a set of $L = 2^b$ discrete values: $x_i \mapsto \hat{x}_{L,i}$, where b is the number of bits in the quantization scheme. This transfer function involves $L-1$ thresholds, which partition $\mathbb{R}$ into L subsets, and L respective weights for these subsets.

4.2. *Quantized Correlation via Sample Covariance*

The traditional correlation estimator for quantized signals matches the form of the continuous covariance estimator, r_∞, to the quantized signals $\hat{r}_L \equiv \langle \hat{x}_L \hat{y}_L\rangle_N$ (Van Vleck & Middleton 1966; Cole 1968; Cooper 1970; Hagen & Farley 1973). In some cases, this result is then appropriately transformed to account for bias. This correlation strategy is optimized through the particular thresholds and weights that determine the transfer function of the quantization.

4.3. *The MLE of Correlation for Quantized Signals*

We now derive the MLE of correlation $r_{L,\mathrm{ML}}$ for quantized signals, which is both consistent and asymptotically efficient. In particular, the likelihood function $\mathcal{L}$ for a set of N independent and identically distributed (i.i.d.) pairs of samples $\{\hat{x}_{L,i}, \hat{y}_{L,i}\}$ drawn from a bivariate normal distribution and then quantized in a scheme with L levels is

$$\mathcal{L}(\rho,\sigma|\{\hat{x}_{L,i},\hat{y}_{L,i}\}) = \prod_{i=1}^{N} P(\hat{x}_{L,i},\hat{y}_{L,i};\rho,\sigma) \qquad (6)$$
$$\Rightarrow \ln \mathcal{L}(\rho,\sigma|\{\hat{x}_{L,i},\hat{y}_{L,i}\}) = \sum_{\ell} \mathcal{N}_\ell \ln \mathcal{P}_\ell(\rho,\sigma).$$

In this expression, ℓ runs over the L^2 possible quantized pairs $\{\hat{x}_L, \hat{y}_L\}$; $\mathcal{N}_\ell$ is the total number of samples in each such category; and $\mathcal{P}_\ell(\rho,\sigma)$ corresponds to the probability of a sampled pair $\{\hat{x}_L, \hat{y}_L\}$ falling in that category.

To determine the MLE, this log-likelihood must be maximized with respect to ρ, if σ is assumed to be known, or with respect to ρ and σ, if σ is unknown. Although we have assumed symmetry $\sigma_{\mathrm{x}} = \sigma_{\mathrm{y}}$, the generalization is straightforward.

The MLE thus requires an evaluation of each probability $\mathcal{P}_\ell$:

$$\mathcal{P}_\ell = S_\ell \int_{R_\ell} dxdy\, P(x,y;\rho,\sigma). \qquad (7)$$

In this expression, $P(x,y;\rho,\sigma)$ is given by Eq. 2, $R_\ell \subset \mathbb{R}^2$ corresponds to the set of unquantized values that map to each quantized state, and $S_\ell \in \mathbb{Z}$ is an optional symmetry factor, to account for the symmetry under inversion, $P(\hat{x}_{L,i},\hat{y}_{L,i}) = P(-\hat{x}_{L,i},-\hat{y}_{L,i})$, and transposition, $P(\hat{x}_{L,i},\hat{y}_{L,i}) = P(\hat{y}_{L,i},\hat{x}_{L,i})$. In a few instances, such as the quadrant integrals that arise in one-bit correlation, Eq. 7 has a simple, closed-form representation. More generally, it can be reduced to a one-dimensional integral of an error function.

Thus, in most cases, the MLE requires minimization over a function that involves one-dimensional numerical integration. However, many strategies can simplify this estimation. For example, if both the number of samples N and quantization bits b are small, then all required solutions can be tabulated. After including the symmetry reductions, the number of distinct correlation possibilities is $N_\ell = 2^{b-1}\left(1+2^{b-1}\right)$. The total number M of partitions of N samples into these categories is then $M = \binom{N+N_\ell-1}{N_\ell-1} \sim N^{N_\ell-1}/(N_\ell-1)!$. If M is prohibitively large, then the N samples can first be partitioned and then the respective correlation estimates averaged to obtain an approximation of the MLE.

4.4. *Noise in the MLE and the Cramér-Rao Lower-Bound*

As we have already mentioned, the CRB determines the minimum variance that any unbiased estimator of ρ can achieve. In terms of the likelihood function of §4.3, the elements of the 2×2 Fisher information matrix for $\{\rho,\sigma\}$ can be written

$$\mathcal{I}_{1,1} \equiv \left\langle \left(\frac{\partial}{\partial\rho} \sum_\ell \mathcal{N}_\ell \ln \mathcal{P}_\ell \right)^2 \right\rangle = N \sum_\ell \frac{\left(\frac{\partial \mathcal{P}_\ell}{\partial\rho}\right)^2}{\mathcal{P}_\ell}, \qquad (8)$$
$$\mathcal{I}_{1,2} = N \sum_\ell \frac{\left(\frac{\partial \mathcal{P}_\ell}{\partial\rho}\right)\left(\frac{\partial \mathcal{P}_\ell}{\partial\sigma}\right)}{\mathcal{P}_\ell},$$
$$\mathcal{I}_{2,2} = N \sum_\ell \frac{\left(\frac{\partial \mathcal{P}_\ell}{\partial\sigma}\right)^2}{\mathcal{P}_\ell}.$$

If σ is known, then the minimum variance of an unbiased estimator of ρ is $\delta r^2_{L,\mathrm{CR}} = \mathcal{I}^{-1}_{1,1}$; if σ is unknown, then the minimum variance is $\delta r^2_{L,\mathrm{CR}} = \mathcal{I}_{2,2}/\left(\mathcal{I}_{1,1}\mathcal{I}_{2,2} - \mathcal{I}^2_{1,2}\right)$. The MLE is asymptotically efficient, so $\langle \delta r^2_{L,\mathrm{ML}} \rangle \to \delta r^2_{L,\mathrm{CR}}$ as $N \to \infty$.

5. EXAMPLES

5.1. *One-bit Quantization*

In the standard one-bit, or two-level, quantization scheme, each sample is reduced to one "sign" bit: $x \mapsto \hat{x}_2 \equiv \mathrm{sign}(x)$. Because the sample error for the signal variance incurs the bulk of the noise in r_∞, quantization actually *improves* upon the estimate of r_∞ in some cases.

Explicitly, we have $r_2 \equiv \langle \hat{x}\hat{y} \rangle_N$. However, this estimate is biased: $\langle r_2 \rangle = 2\pi^{-1} \sin^{-1}\rho$. The standard Van Vleck clipping correction, denoted $r_{2,\mathrm{V}}$, improves the bias to $\mathcal{O}(1/N)$ by simply inverting this relationship (Van Vleck & Middleton 1966):

$$r_{2,\mathrm{V}} \equiv \sin\left(\frac{\pi}{2} r_2\right). \qquad (9)$$

In fact, $r_{2,\mathrm{V}}$ gives precisely the MLE. To see this, note that the quantized products, $\hat{x}\hat{y}$, have probability $P(\pm 1) = \frac{1}{2} \pm \frac{1}{\pi}\arcsin\rho$. Minimizing the log-likelihood (Eq. 6) with respect to ρ gives that $r_{2,\mathrm{ML}} = r_{2,\mathrm{V}}$.

Because $r_{2,\mathrm{V}}$ is the MLE, the noise for large N is given by the CRB. Substituting $P(\pm 1)$ into Eq. 8 gives

$$N\delta r^2_{2,\mathrm{CR}} = \left[\left(\frac{\pi}{2}\right)^2 - (\arcsin\rho)^2\right]\left(1-\rho^2\right). \qquad (10)$$

We can easily verify that the noise in $r_{2,\mathrm{V}}$ actually achieves this lower bound. Namely, the correlation estimate r_2 is a one-dimensional random walk with N steps of length $\pm 1/N$, distributed according to $P(\pm 1)$. For large N, the central limit theorem gives that r_2 follows a Gaussian distribution with mean $2\pi^{-1}\arcsin\rho$ and variance $N^{-1}\left[1 - \left(2\pi^{-1}\arcsin\rho\right)^2\right]$. In this limit, we obtain

$$\langle r^2_{2,\mathrm{V}} \rangle = \frac{1}{2}\left\{1 - \left(1-2\rho^2\right)\exp\left[\frac{4(\arcsin\rho)^2 - \pi^2}{2N}\right]\right\}$$
$$\Rightarrow N\langle \delta r^2_{2,\mathrm{V}} \rangle \approx \left[\left(\frac{\pi}{2}\right)^2 - (\arcsin\rho)^2\right]\left(1-\rho^2\right), \qquad (11)$$

which is identical to the CRB.

The most striking improvement of $r_{2,\mathrm{V}}$ relative to r_∞ occurs as $\rho \to \pm 1$; in this limit, the one-bit correlation has no noise, while $\langle \delta r^2_\infty \rangle = 2/N$.

5.2. *Two-bit Quantization*

Perhaps the most common quantization strategy replaces each sample by a pair of bits for sign and magnitude. The (non-zero) thresholds $\pm v_0$ are fixed at some level relative to the estimated RMS signal voltage σ in a way that minimizes the expected RMS noise in the subsequent correlation estimates. The resulting four levels are then assigned weights $\hat{x}_2 \in \{\pm 1, \pm n\}$, where n is also chosen to minimize the noise. In terms of the mean quantized product $r_4 \equiv \langle \hat{x}_4 \hat{y}_4 \rangle_N$, one obtains the correlation estimate (TMS; Eq. 8.43)

$$r_{4,\mathrm{v}} = \frac{r_4}{\Phi + n^2(1-\Phi)}, \qquad \Phi \equiv \mathrm{erf}\left(\frac{v_0}{\sigma\sqrt{2}}\right). \qquad (12)$$

This estimate of correlation, which already assumes exact knowledge of σ, retains a significant ($\sim$10%) bias to high $|\rho|$ (see Figure 1 of Jenet & Anderson (1998)). If $|\rho| \lesssim 0.8$, for instance, then the appropriate correction is simply a constant scaling factor (TMS; Eq. 8.45):

$$r_{4,\mathrm{V}} \approx \left\{ \frac{\pi\left[\Phi + n^2(1-\Phi)\right]}{2\left[(n-1)\,E + 1\right]^2} \right\} r_{4,\mathrm{v}}, \qquad E \equiv e^{-\frac{1}{2}\left(\frac{v_0}{\sigma}\right)^2}. \qquad (13)$$

For additional details and a complete formulation to remove the bias, see Gwinn (2004). Here, we use the "V" subscript to draw analogy with the Van Vleck correction for one-bit correlation. Namely, this estimate calculates the sample covariance after quantization and then performs a bias correction according to the estimated correlation. The remaining bias is $\mathcal{O}(1/N)$.

Researchers then optimize this two-bit correlation scheme by a particular choice of thresholds and weights: $v_0 = 0.9815$, $n = 3.3359$. However, unlike one-bit correlation, the bias-corrected quantized product $r_{4,\mathrm{V}}$ is *not* the optimal estimator of correlation for quantized data. In particular, $r_{4,\mathrm{V}}$ even reflects the disturbing feature of the continuous estimate r_∞ that the noise tends to increase with $|\rho|$. Hence, high correlations present the paradoxical situation in which traditional estimates $\{r_\infty, r_{4,\mathrm{V}}, r_{2,\mathrm{V}}\}$ perform *better* as the number of bits is *reduced*.

This troubling evolution merely reflects the incompleteness of these correlation estimates. Figure 2 compares the noise in $r_{4,\mathrm{V}}$ to the noise in the MLE, both when σ is known and unknown. Each MLE has negligible bias and noise that reflects the behavior seen in the corresponding unquantized MLE, r_p or r_q. The noise is always lower than that of $r_{2,\mathrm{V}}$ and approaches zero as $|\rho| \to 1$. We therefore resolve the puzzling evolution of correlation noise after quantization.

The only remaining barrier is the computational difficulty of implementation. However, for small values of N, the maximum-likelihood solutions may be tabulated prior to calculation; the required number of tabulated values is $M = \binom{N+5}{5} \sim N^5/5!$ (see §4.3). Alternatively, one can first partition the N samples, then calculate the MLE of correlation for each subset via tabulation, and finally average the results. We defer a comprehensive treatment of these implementation strategies to a future work.

5.3. *Many-bit Quantization*

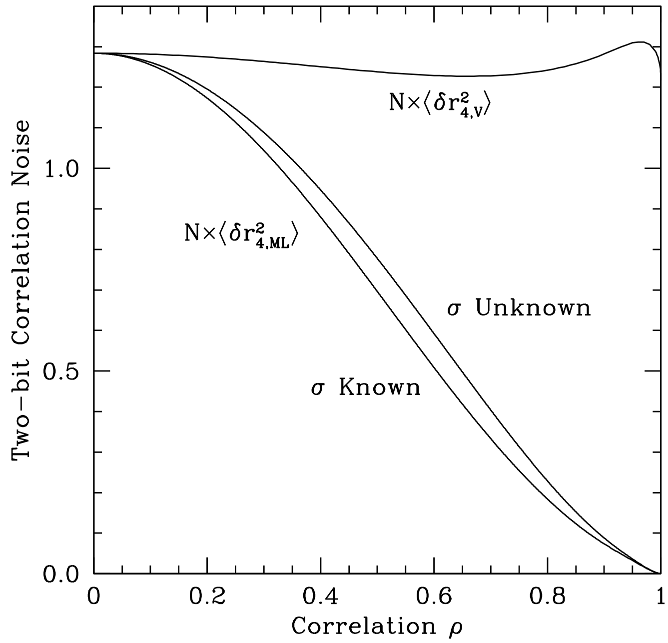

Figure 2. Noise in estimates of correlation for signals quantized with two bits. The chosen levels ($v_0 = 0.9815$) and weights ($n = 3.3359$) are optimal as $|\rho| \to 0$. The upper curve gives the noise in the traditional estimator via sample covariance, as derived in Gwinn (2004), whereas the lower curves give the noise in the MLEs with and without knowledge of σ.

Modern instrumentation now permits the storage of baseband data with many-bit quantization schemes. In this case, the noise in the MLE of correlation rapidly approaches that in the corresponding unquantized limit, r_p or r_q (see Figure 3). In such cases, these estimators for unquantized signals provide excellent approximations of the quantized MLEs, and the primary concerns are the influence of RFI and instrumental limitations (TMS). Furthermore, although low-bit quantization schemes are quite robust to impulsive RFI, estimates such as r_p are not, so alternative quantization schemes that are robust at the expense of increased noise may be preferred (Fridman 2009).

We now consider the incurred bias when approximating the quantized MLE by r_p. Specifically, consider a high-bit scheme with L levels, thresholds in multiples of $\pm v_0$, and quantization weights $\hat{x}$ that are the average values of their respective preimages. We denote the corresponding estimator $r_{L,\mathrm{p}}$. Then, if the highest thresholds extend far into the tail of the distribution, the bias after quantization is approximately

$$\delta r_{L,\mathrm{p}} \approx -\frac{1}{12}\left(\frac{v_0}{\sigma}\right)^2 \rho. \qquad (14)$$

For more general expressions, which include the effects of the finite outer thresholds, consult the discussion in §8.3 of TMS. While correcting the bias is straightforward, even for a low number of bits, this strategy is ineffective for low-bit schemes because $r_{L,\mathrm{p}}$ is not a sufficient statistic for ρ.

6. REDUCTIONS FOR SMALL CORRELATION

Although the MLE of correlation decreases the noise for large $|\rho|$, it exhibits identical noise to traditional estimators at small $|\rho|$. Furthermore, in this limit, knowl-

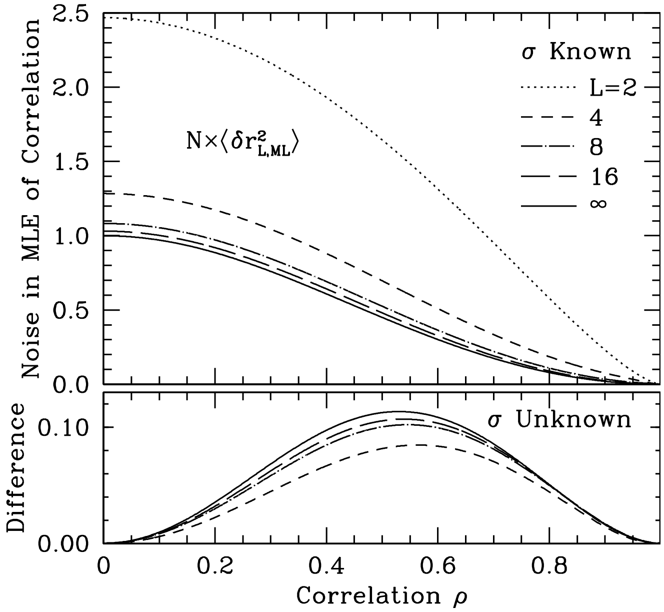

Figure 3. Noise in the MLE of correlation for quantized signals as a function of correlation for various quantization levels. The upper panel shows the noise when σ is known, and the lower panel shows the additional noise when σ is unknown. For simplicity, we set the $(L-1)$ quantization thresholds in multiples of $\pm 4/L$. Observe that the reduction in noise provided by knowledge of σ becomes more pronounced as the number of levels is *increased*. For example, knowledge of σ provides no improvement for one-bit correlation.

edge of σ does not reduce the noise. These features both arise from the form of the bivariate Gaussian PDF in this limit.

For example, consider the estimation of correlation for unquantized signals when $\sigma_{\rm x} = \sigma_{\rm y} = 1$ is known. Then, if $|\rho| \ll 1$, the joint PDF of N independently-drawn pairs of correlated random variables $\{x_i, y_i\}$ is (see Eq. 3)

$$P(\{x_i, y_i\}; \rho) \approx \frac{1}{(2\pi)^N} \left(1 + N\rho\langle xy\rangle_N\right) e^{-\frac{N}{2}\left(\langle x^2\rangle_N + \langle y^2\rangle_N\right)}. \tag{15}$$

Thus, from the factorization criterion, $r_\infty \equiv \langle xy\rangle_N$ is a sufficient statistic for ρ, and so we expect the asymptotic noise in r_∞ to match that of $r_{\rm q}$ as $\rho \to 0$.

Likewise, consider the joint distribution of the samples after quantization into L weighted levels. In this case, we require the set of quantized probabilities

$$\mathcal{P}_\ell \approx \frac{1}{2\pi} \int_{R_\ell} dxdy\ (1 + \rho xy)\, e^{-\frac{1}{2}\left(x^2+y^2\right)} \equiv \alpha_\ell + \beta_\ell \rho. \tag{16}$$

The joint PDF of the quantized samples is then

$$\begin{aligned} P(\{\hat{x}_{L,i}, \hat{y}_{L,i}\}; \rho) &= \prod_\ell \mathcal{P}_\ell^{\mathcal{N}_\ell} \qquad (17) \\ &\approx \left(\prod_\ell \alpha_\ell^{\mathcal{N}_\ell}\right) \left(1 + \rho \sum_\ell \mathcal{N}_\ell \frac{\beta_\ell}{\alpha_\ell}\right). \end{aligned}$$

Hence, for small $|\rho|$, the factorization criterion gives that $\langle w(\hat{x}, \hat{y})\rangle$ is a sufficient statistic for ρ, if the weight function is determined by

$$\begin{aligned} w(\hat{x}, \hat{y}) &= \frac{\int_{R_\ell} dxdy\, xy e^{-\frac{1}{2}\left(x^2+y^2\right)}}{\int_{R_\ell} dxdy\, e^{-\frac{1}{2}(x^2+y^2)}} \qquad (18) \\ &= \left[\frac{\int_{R_{\ell,x}} dx\, x e^{-\frac{x^2}{2}}}{\int_{R_{\ell,x}} dx\, e^{-\frac{x^2}{2}}}\right] \left[\frac{\int_{R_{\ell,y}} dy\, y e^{-\frac{y^2}{2}}}{\int_{R_{\ell,y}} dy\, e^{-\frac{y^2}{2}}}\right], \end{aligned}$$

where $R_{\ell,x} \subset \mathbb{R}$ defines the range of values spanned by each quantized level.

Moreover, the final factorization in Eq. 18 demonstrates that, by assigning an appropriate weight to each quantization level: $w(\hat{x}, \hat{y}) = \hat{x}\hat{y}$, the sample covariance is a sufficient statistic for ρ and will achieve optimal noise performance as $|\rho| \to 0$.

The asymptotic noise in this limit is then the CRB:

$$\delta r^2_{L,\rm CR}\big\rfloor_{\rho=0} = \frac{2\pi}{N} \left\{ \sum_\ell \frac{\left[\int_{R_\ell} dxdy\, xy e^{-\frac{1}{2}\left(x^2+y^2\right)}\right]^2}{\int_{R_\ell} dxdy\, e^{-\frac{1}{2}(x^2+y^2)}} \right\}^{-1}. \tag{19}$$

Minimizing this equation yields the optimal thresholds. Then, Eq. 18 immediately determines the optimal weights. Observe that these weights are slightly different than those of some previous works, such as Jenet & Anderson (1998), but match the ratios of traditional quantization schemes, such as $n = 3.336$ when $v_0 = 0.982$ for two-bit correlation, for instance.

Finally, $\mathcal{I}_{1,2} \to 0$ as $\rho \to 0$. This result follows easily by substituting $\mathcal{P}_\ell$ and its derivatives into Eq. 8. Hence, the CRB is unchanged by knowledge of σ in this limit.

7. SUMMARY

We have explored the paradoxical scaling of noise in traditional estimates of correlation for quantized signals. In particular, we have shown that the decrease in noise that quantization affords is a result of an incomplete comparison with unquantized correlation schemes and of suboptimal correlation strategies for quantized signals.

We have derived the MLE of correlation, both with and without knowledge of the signal variance and quantization, and we have compared these estimates to traditional schemes. The MLE has negligible bias, lower noise, and is asymptotically efficient: for a large number of samples, no other unbiased scheme will achieve lower noise. We have also derived simple expressions for this asymptotic noise (the CRB). While the MLE gives the familiar Van-Vleck corrected sample covariance for one-bit quantization, it differs significantly from current correlation strategies for all other cases.

Nevertheless, traditional correlation schemes are fully optimized in the limit $\rho \to 0$. Namely, for suitably chosen weights, the sample covariance $\hat{r}_L$ is a sufficient statistic for the correlation ρ, in this limit.

Future detectors, such as the Square-Kilometer Array, that will achieve high signal-to-noise while being limited to a small number of quantization bits, can benefit from these novel correlation strategies to reduce both the distortion and noise introduced by quantization.

We thank the U.S. National Science Foundation for

financial support for this work (AST-1008865). The work of H. C. was supported by a UCSB SURF award.

APPENDIX

A. MLE FOR UNQUANTIZED SIGNALS WITH KNOWN VARIANCE

We now summarize the main features of the MLE of correlation for samples $\{x_i, y_i\}$ drawn from a bivariate Gaussian distribution with known means and variances. See Kendall & Stuart (1979) for additional details. For simplicity, we assume that the means are zero and the variances are unity. We also assume that each pair is drawn independently. The likelihood function is then

$$\mathcal{L}(\rho|\{x_i, y_i\}) \equiv \prod_{i=1}^{N} P(x_i, y_i; \rho, \sigma_{\rm x}, \sigma_{\rm y}) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2\left(1-\rho^2\right)} \sum_{i=1}^{N} \left(x_i^2 + y_i^2 - 2\rho x_i y_i\right)\right]. \tag{A1}$$

The condition for the likelihood function to be extremized is

$$r_\infty \left(1+\rho^2\right) - \rho\left(s_{\rm x}^2 + s_{\rm y}^2 - \left(1-\rho^2\right)\right) = 0, \tag{A2}$$

where $s_{\rm x}^2 \equiv \langle x^2 \rangle_N$, $s_{\rm y}^2 \equiv \langle y^2 \rangle_N$, and $r_\infty \equiv \langle xy \rangle_N$. Hence, the triplet $\{r_\infty, s_{\rm x}, s_{\rm y}\}$ is sufficient for ρ. We will denote the appropriate solution to this cubic equation $r_{\rm q}$.

To obtain some intuition for this result, let $\epsilon \equiv s_{\rm x}^2 + s_{\rm y}^2 - 2$. Then $\langle \epsilon^2 \rangle = 4(1+\rho^2)/N$. The discriminant of the cubic is

$$\Delta = -4r_\infty^4 + \left(\epsilon^2 + 20\epsilon - 8\right) r_\infty^2 - 4\left(1+\epsilon\right)^3. \tag{A3}$$

If $\Delta < 0$, then the cubic has a single real solution. As a rough rule of thumb, we can simply consider when all terms are negative. Since $\delta\epsilon \approx 2/\sqrt{N}$, we see that there is likely a unique real solution whenever $\epsilon < .39$, or $N \gtrsim 25$.

Although finding this solution is both analytically and numerically straightforward, an approximation is both useful and enlightening:

$$r_{\rm q} = r_\infty \left[1 - \frac{1}{1+r_\infty^2}\epsilon + \frac{1-r_\infty^2}{\left(1+r_\infty^2\right)^3}\epsilon^2 + \mathcal{O}\left(\epsilon^3\right)\right]. \tag{A4}$$

This expansion immediately identifies the appropriate root of the cubic equation. Furthermore, we can determine the asymptotic noise for $r_{\rm q}$ by expanding Eq. A4 for large N:

$$\left\langle \delta r_{\rm q}^2 \right\rangle = \left\langle \delta r_\infty^2 \right\rangle + \frac{\rho^2}{\left(1+\rho^2\right)^2}\left\langle \epsilon^2 \right\rangle - \frac{2\rho}{1+\rho^2}\left\langle \delta r_\infty\ \epsilon \right\rangle. \tag{A5}$$

A straightforward application of Isserlis' Theorem (Isserlis 1918) gives that $\left\langle \delta r_\infty^2 \right\rangle = (1+\rho^2)/N$, $\left\langle \epsilon^2 \right\rangle = 4\left(1+\rho^2\right)/N$, and $\langle \delta r_\infty \epsilon \rangle = 4\rho/N$. Putting everything together, we obtain

$$\lim_{N\to\infty} N \left\langle \delta r_{\rm q}^2 \right\rangle = \frac{\left(1-\rho^2\right)^2}{1+\rho^2}. \tag{A6}$$

We can easily verify that this result is equal to the CRB:

$$\begin{aligned} \left\langle \delta r_{\rm CR}^2 \right\rangle &= \left\{ N \int_{-\infty}^{\infty} dxdy\, \frac{\left(\frac{\partial P(x,y;\rho)}{\partial \rho}\right)^2}{P(x,y;\rho)} \right\}^{-1} \qquad \text{(A7)} \\ &= \left\{ \frac{N}{\left(1-\rho^2\right)^4} \int_{-\infty}^{\infty} dxdy\, \left[-\rho\left(1-\rho^2\right) - \left(1+\rho^2\right)xy + \rho\left(x^2+y^2\right)\right]^2 \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{\left(x^2+y^2-2\rho xy\right)}{2\left(1-\rho^2\right)}\right] \right\}^{-1} \\ &= \frac{1}{N}\frac{\left(1-\rho^2\right)^2}{1+\rho^2}. \end{aligned}$$

REFERENCES

Anderson, T. W. 1996, Statistical Science, 11, pp. 20
Cole, T. 1968, Australian Journal of Physics, 21, 273
Cooper, B. F. C. 1970, Australian Journal of Physics, 23, 521
Deller, A. T., Tingay, S. J., Bailes, M., & West, C. 2007, PASP, 119, 318
Dewdney, P. E., Hall, P. J., Schilizzi, R. T., & Lazio, T. J. L. W. 2009, IEEE Proceedings, 97, 1482
Fisher, R. A. 1915, Biometrika, 10, pp. 507
—. 1921, Metron, 1, 3

Fridman, P. 2009, A&A, 502, 401
Gwinn, C. 2004, PASP, 116, 84
Gwinn, C. R. 2006, PASP, 118, 461
Hagen, J. B., & Farley, D. T. 1973, Radio Science, 8, 775
Hotelling, H. 1953, Journal of the Royal Statistical Society. Series B (Methodological), 15, pp. 193
Isserlis, L. 1918, Biometrika, 12, pp. 134
Jenet, F. A., & Anderson, S. B. 1998, PASP, 110, 1467
Kendall, M., & Stuart, A. 1979, The advanced theory of statistics. Vol.2: Inference and relationship
Lehmann, E. L., & Casella, G. 1998, Theory of Point Estimation (New York: Springer-Verlag)
Thompson, A. R., Moran, J. M., & Swenson, Jr., G. W. 2001, Interferometry and Synthesis in Radio Astronomy, 2nd Edition
Van Vleck, J. H., & Middleton, D. 1966, Proceedings of the IEEE, 54, 2